Модели информационного поиска

Лекция 2

БГУ ФПМИ, 2018

План

Булев поиск
- Инвертированный индекс

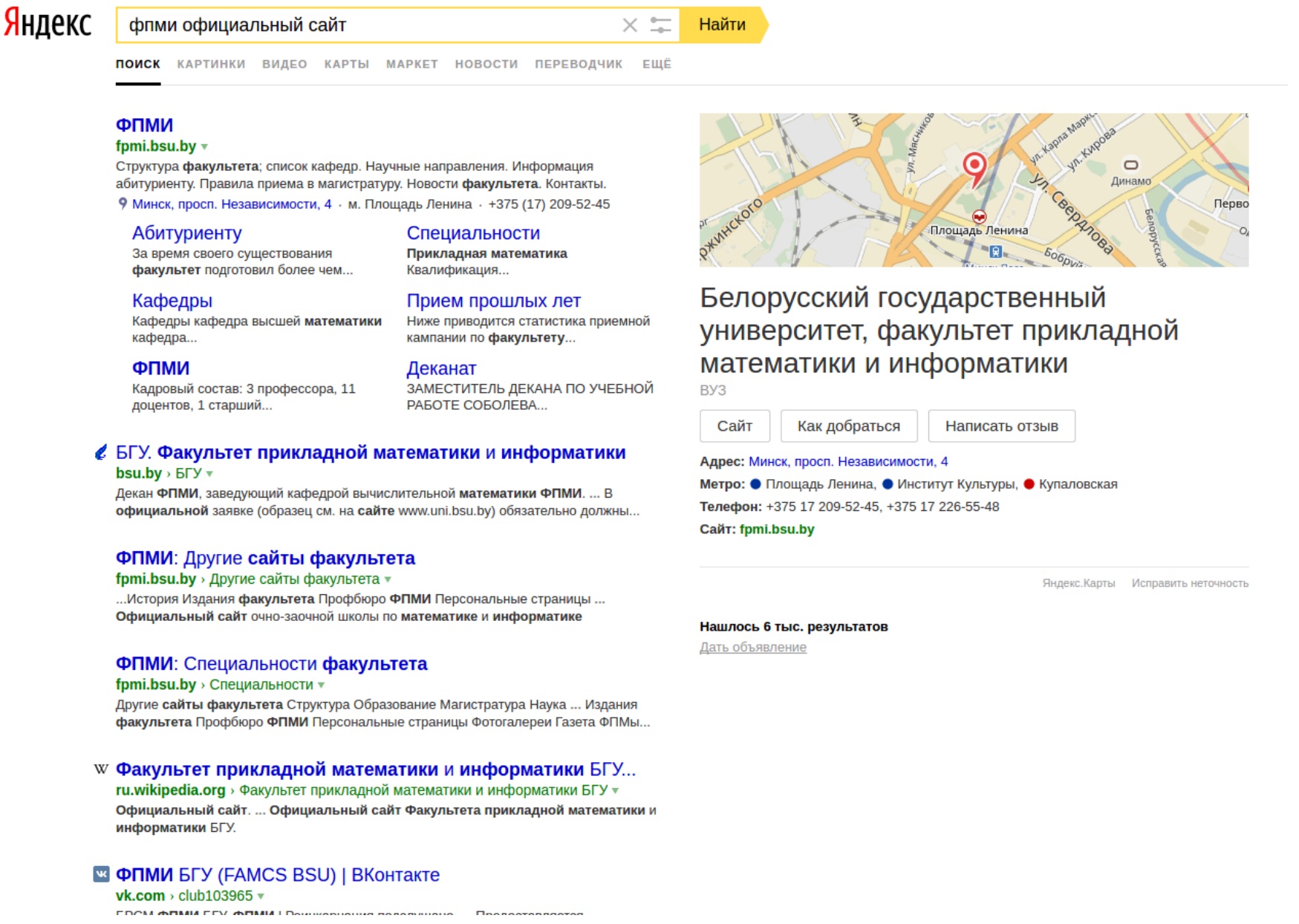
Векторная модель

Вероятностные модели в информационном поиске
- Языковые модели

Сочетание признаков

Релевантность

- Сложное понятие, зависящее от субъективного восприятия.
- В рамках моделей информационного поиска рассматривается с нескольких сторон:
  - тематическая релевантность
  - пользовательская релевантность
  - бинарная релевантность
  - многозначная релевантность


Яндекс
фпми официальный сайт
Найти
ПОИСК КАРТИНКИ ВИДЕО КАРТЫ МАРКЕТ НОВОСТИ ПЕРЕВОДЧИК ЕЩЁ
ФПМИ
fpmi.bsu.by
Структура факультета; список кафедр. Научные направления. Информация абитуриенту. Правила приема в магистратуру. Новости факультета. Контакты.
Минск, просп. Независимости, 4 · м. Площадь Ленина · +375 (17) 209-52-45
Абитуриенту
За время своего существования факультет подготовил более чем...
Специальности
Прикладная математика
Квалификация...
Кафедры
Кафедры кафедра высшей математики кафедра...
Прием прошлых лет
Ниже приводится статистика приемной кампании по факультету...
ФПМИ
Кадровый состав: 3 профессора, 11 доцентов, 1 старший...
Деканат
ЗАМЕСТИТЕЛЬ ДЕКАНА ПО УЧЕБНОЙ РАБОТЕ СОБОЛЕВА...
БГУ. Факультет прикладной математики и информатики
bsu.by › БГУ
Декан ФПМИ, заведующий кафедрой вычислительной математики ФПМИ. ... В официальной заявке (образец см. на сайте www.uni.bsu.by) обязательно должны...
ФПМИ: Другие сайты факультета
fpmi.bsu.by › Другие сайты факультета
...История Издания факультета Профбюро ФПМИ Персональные страницы ... Официальный сайт очно-заочной школы по математике и информатике
ФПМИ: Специальности факультета
fpmi.bsu.by › Специальности
Другие сайты факультета Структура Образование Магистратура Наука ... Издания факультета Профбюро ФПМИ Персональные страницы Фотогалереи Газета ФПМы...
Факультет прикладной математики и информатики БГУ...
ru.wikipedia.org › Факультет прикладной математики и информатики БГУ
Официальный сайт. ... Официальный сайт Факультета прикладной математики и информатики БГУ.
ФПМИ БГУ (FAMCS BSU) | ВКонтакте
vk.com › club103965
ул. Мясникова
ул. Карла Маркса
ул. Кирова
ул. Свердлова
Динамо
Площадь Ленина
Белорусская
Белорусский государственный университет, факультет прикладной математики и информатики
ВУЗ
Сайт
Как добраться
Написать отзыв
Адрес: Минск, просп. Независимости, 4
Метро: Площадь Ленина, Институт Культуры, Купаловская
Телефон: +375 17 209-52-45, +375 17 226-55-48
Сайт: fpmi.bsu.by
Яндекс.Карты
Исправить неточность
Нашлось 6 тыс. результатов
Дать объявление

Яндекс | могадишо | Найти

ПОИСК КАРТИНКИ ВИДЕО КАРТЫ МАРКЕТ НОВОСТИ ПЕРЕВОДЧИК ЕЩЁ

Могадишо — Википедия
ru.wikipedia.org › Могадишо
Могади́шо (сомал. Muqdisho, араб. مقديش, итал. **Mogadiscio**) — столица Сомали, крупнейший город и главный порт страны, являющийся также её культурным, финансовым и индустриальным центром.

Могадишо на карте Сомали
yandex.ru › Могадишо
Могадишо на карте Сомали — схематической или спутниковой. Поиск на карте по адресу или названию населённого пункта.

?

могадишо — смотрите картинки
yandex.by/images › могадишо
Пожаловаться

??

Могадишо — новости

В **Могадишо** в результате теракта 18 человек...
wordyou.ru 20 фев 2017
В **Могадишо** в результате теракта 18 человек погибли и 25 пострадали » "Слово без границ" - новости России и мира сегодня.

Число жертв взрыва на рынке в **Могадишо**...
belta.by 20 фев 2017

В результате взрыва в столице Сомали **Могадишо**...
vladtime.ru 19 фев 2017

Могадишо, Сомали - отдых, погода, отзывы... | RestBee.ru
restbee.ru › world/afrika/somali/mogadisho
Могадишо является не только официальной столицей Сомали, но и крупнейшим

Могадишо

Столица Сомали

Столица Сомали, крупнейший город и главный порт страны, являющийся также её культурным, финансовым и индустриальным центром. Население города - 2 120 000 жителей. Площадь - 91 км². Википедия

Погода: 27°C, Ясно
Местное время: 22 февраля, 00:30
Дата возникновения: 1331 г.
Население: 2 120 000 чел.
Площадь: 91 км²

Смотрите также

Сомали | Кисмайо | Мапуту | Аддис-Абеба | Дакар | Хартум

Википедия Сообщить об ошибке

Яндекс

джанго освобожденный

Найти

ПОИСК КАРТИНКИ ВИДЕО КАРТЫ МАРКЕТ НОВОСТИ ПЕРЕВОДЧИК ЕЩЁ

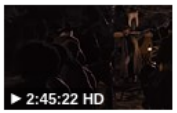
Джанго освобожденный (2012) смотреть онлайн...
kinogo.club › Джанго освобожденный
Квентин Тарантино - это **режиссер**, создавший лучший вестерн всех времен "**Джанго освобожденный**".

Джанго освобожденный (2012) — КиноПоиск
Кадры Промо Съёмки Скриншоты Фан-арт Обложки Концепт
kinopoisk.ru › Джанго
Драма, вестерн, приключения. Режиссер: Квентин Тарантино. В ролях: Джейми Фокс, Кристоф Вальц, Леонардо ДиКаприо и др. Эксцентричный охотник за головами, также известный как «Дантист», промышляет отстрелом самых опасных преступников.
★★★★☆ 8,2/10 · 268 663 оценок

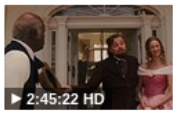
Джанго освобождённый — Википедия
ru.wikipedia.org › Джанго освобождённый
«**Джанго освобождённый**» (англ. **Django** Unchained) — художественный фильм 2012 года режиссёра Квентина Тарантино в жанре спагетти-вестерн.

Джанго освобожденный (2013) смотреть онлайн фильм...
kinokrad.co › Джанго освобожденный
В прокат вышел блокбастер «**Джанго освобождённый**» (2013), в котором режиссеру удалось задействовать лучших актёров Голливуда.

джанго освобожденный — 10 тыс. видео
video.yandex.by › джанго освобожденный



Джанго освобожденный / 2012 / Фильм / Full...
video.mail.ru



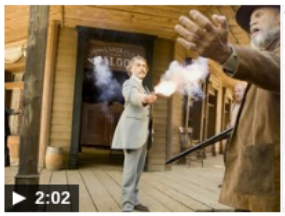
Джанго освобожденный
ok.ru

53. **Джанго освобождённый**
video.mail.ru

Джанго
ok.ru

Джанго Освобожденный

Django Unchained, 2012 18+



Эксцентричный охотник за головами, также известный как «Дантист», промышляет отстрелом самых опасных преступников. Работенка пыльная, и без надежного помощника ему не обойтись. Но как найти такого и желательно не очень дорогого? Беглый раб по имени… Читать дальше

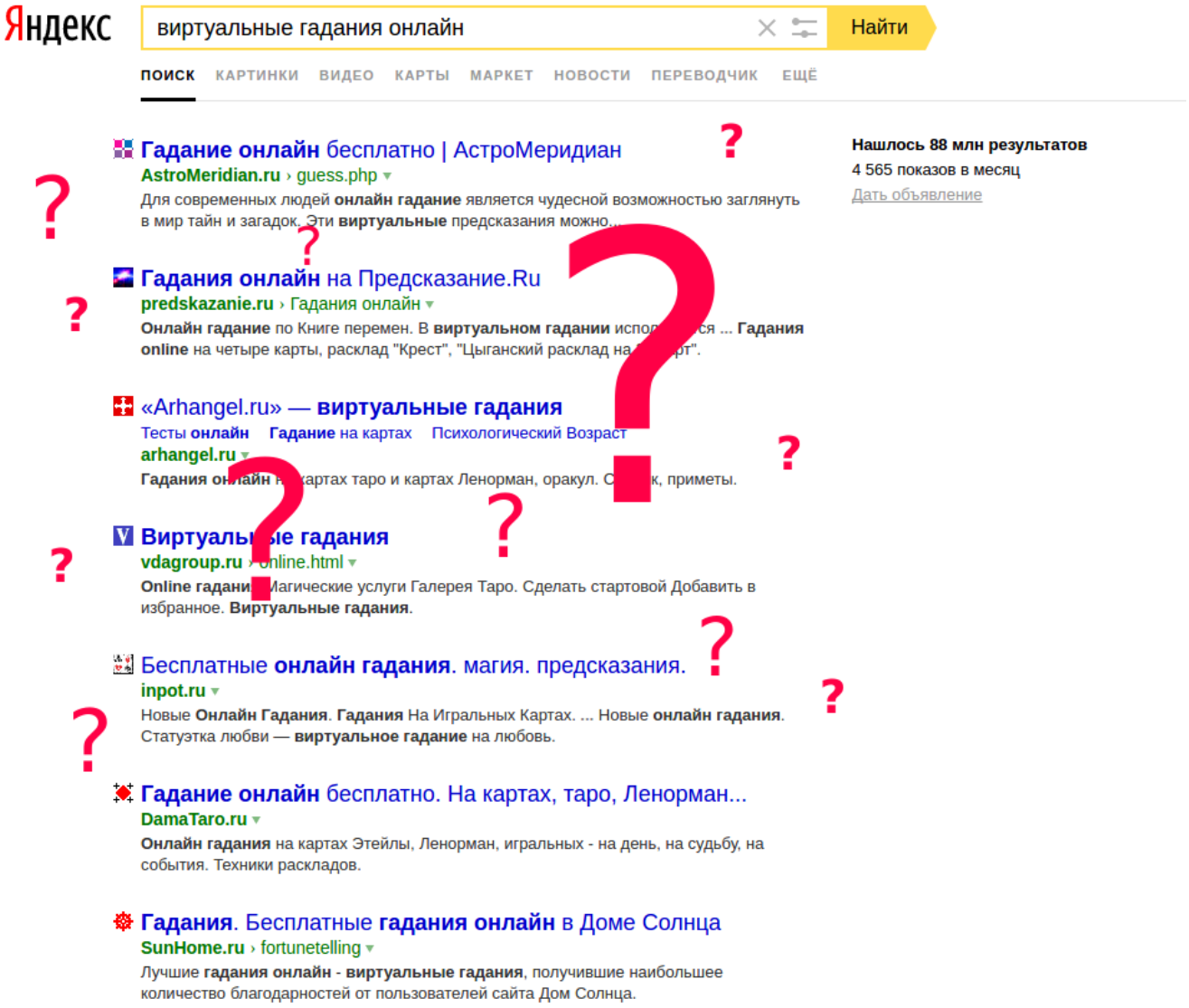
★★★★☆ 8,2/10 КиноПоиск
★★★★½ 8,5/10 IMDb

Жанр: драма, Вестерн, приключения, комедия
Страна: США
Режиссёр: Квентин Тарантино
Музыка: Эннио Морриконе
Длительность: 141 мин.
Продюсеры: Стейси Шер, Боб Вайнштейн, Харви Вайнштейн, Реджинальд Хадлин, Майкл Шамберг, Уильям Пол Кларк, Джеймс В. Скотчдопоул

Актеры







Яндекс
виртуальные гадания онлайн
Найти
ПОИСК КАРТИНКИ ВИДЕО КАРТЫ МАРКЕТ НОВОСТИ ПЕРЕВОДЧИК ЕЩЁ
Гадание онлайн бесплатно | АстроМеридиан
AstroMeridian.ru › guess.php
Для современных людей онлайн гадание является чудесной возможностью заглянуть в мир тайн и загадок. Эти виртуальные предсказания можно...
Гадания онлайн на Предсказание.Ru
predskazanie.ru › Гадания онлайн
Онлайн гадание по Книге перемен. В виртуальном гадании испо ся ... Гадания online на четыре карты, расклад "Крест", "Цыганский расклад на рт".
«Arhangel.ru» — виртуальные гадания
Тесты онлайн Гадание на картах Психологический Возраст
arhangel.ru
Гадания онлайн на картах таро и картах Ленорман, оракул. С к, приметы.
Виртуальные гадания
vdagroup.ru › online.html
Online гадани Магические услуги Галерея Таро. Сделать стартовой Добавить в избранное. Виртуальные гадания.
Бесплатные онлайн гадания. магия. предсказания.
inpot.ru
Новые Онлайн Гадания. Гадания На Игральных Картах. ... Новые онлайн гадания. Статуэтка любви — виртуальное гадание на любовь.
Гадание онлайн бесплатно. На картах, таро, Ленорман...
DamaTaro.ru
Онлайн гадания на картах Этейлы, Ленорман, игральных - на день, на судьбу, на события. Техники раскладов.
Гадания. Бесплатные гадания онлайн в Доме Солнца
SunHome.ru › fortunetelling
Лучшие гадания онлайн - виртуальные гадания, получившие наибольшее количество благодарностей от пользователей сайта Дом Солнца.
Нашлось 88 млн результатов
4 565 показов в месяц
Дать объявление

Правильная модель информационного поиска позволяет находить релевантные документы по заданному запросу.

План

Булев поиск
Инвертированный индекс

Векторная модель

Вероятностные модели в информационном поиске
Языковые модели

Сочетание признаков

Булева модель поиска

- Запрос имеет вид булева выражения, состоящего из термов и операторов AND, OR, NOT.

  (”семь” OR ”один”) AND NOT ”все”

- Основана на точном совпадении.
- Релевантность бинарна.

Булева модель поиска

Результат поиска – неупорядоченное множество документов, удовлетворяющих запросу.

Булева модель поиска

Результат поиска – неупорядоченное множество документов, удовлетворяющих запросу.

1. "Семь раз отмерь, один раз отрежь."
2. "Один за всех, все за одного."
3. "Семь бед — один ответ."
4. "Семь вёрст до небес и все лесом."

Query = ("семь" OR "один") AND NOT "все"
Result docs = $(\{1, 3, 4\} \cup \{1, 2, 3\}) \cap \overline{\{2, 4\}} = \{\mathbf{1}, \mathbf{3}\}$

Как это работает?

~~Последовательно просмотрим все слова запроса в каждом документе~~.

Как это работает?

~~Последовательно просмотрим все слова запроса в каждом документе~~.

	1	2	3	4
беда	0	0	1	0
верста	0	0	0	1
все	0	1	0	1
один	1	1	1	0
семь	1	0	1	1
...				

$$(\mathbf{1011} \vee \mathbf{1110}) \wedge \neg(\mathbf{0101}) = \mathbf{1010}$$

Словарь		Словопозиции (postings)
Терм	$\mathbf{N_t}$	
беда	3	3, 10, 11
верста	2	4, 5
друг	7	11, 14, 18, 21, 25, 34, 40
семь	10	1, 3, 4, 11, 15, 23, 37, 45, 51, 56
.....		.....

В общем случае

$$\text{posting}(d, t) = (d, f_{t,d}, [p_1, \ldots, p_{f_{t,d}}])$$

Построение инвертированного индекса

1. ['Семь раз отмерь, один раз отрежь.', 'Один за всех, все за одного', 'Семь бед — один ответ.', 'Семь вёрст до небес и все лесом.']

Построение инвертированного индекса

1. ['Семь раз отмерь, один раз отрежь.', 'Один за всех, все за одного', 'Семь бед — один ответ.', 'Семь вёрст до небес и все лесом.']
2. [['семь', 'раз', 'отмерить', 'один', 'раз', 'отрезать'], ['один', 'за', 'все', 'все', 'за', 'один'], ['семь', 'беда', 'один', 'ответ'], ['семь', 'верста', 'до', 'небеса', 'и', 'все', 'лес']]

Построение инвертированного индекса

1. ['Семь раз отмерь, один раз отрежь.', 'Один за всех, все за одного', 'Семь бед — один ответ.', 'Семь вёрст до небес и все лесом.']
2. [['семь', 'раз', 'отмерить', 'один', 'раз', 'отрезать'], ['один', 'за', 'все', 'все', 'за', 'один'], ['семь', 'беда', 'один', 'ответ'], ['семь', 'верста', 'до', 'небеса', 'и', 'все', 'лес']]
3. [('беда', 3), ('верста', 4), ('все', 2), ('все', 2), ('все', 4), ('до', 4), ('за', 2), ('за', 2), ('и', 4), ('лес', 4), ('небеса', 4), ('один', 1), ('один', 2), ('один', 2), ('один', 3), ('ответ', 3), ('отмерить', 1), ('отрезать', 1), ('раз', 1), ('раз', 1), ('семь', 1), ('семь', 3), ('семь', 4)]

Терм	Словопозиции
’беда’	[3]
’верста’	[4]
’все’	[2, 4]
’до’	[4]
’за’	[2]
’и’	[4]
’лес’	[4]
’небеса’	[4]
’один’	[1, 2, 3]
’ответ’	[3]
’отмерить’	[1]
’отрезать’	[1]
’раз’	[1]
’семь’	[1, 3, 4]

Поиск в индексе

- Два списка docId пересекаются аналогично алгоритму merge.

Поиск в индексе

- Два списка docId пересекаются аналогично алгоритму merge.
- Для ускорения работы обработку списков нужно производить в порядке возрастания их длин.

Булева модель, +/−

+

Булева модель, +/−

\+

- предсказуемость;

Булева модель, +/−

\+

- предсказуемость;
- легкая интерпретация результатов;

Булева модель, +/−

\+

- предсказуемость;
- легкая интерпретация результатов;
- эффективность.

−

Булева модель, +/−

\+

- предсказуемость;
- легкая интерпретация результатов;
- эффективность.

−

- качество полностью зависит от пользователя;

Булева модель, +/−

\+

- предсказуемость;
- легкая интерпретация результатов;
- эффективность.

−

- качество полностью зависит от пользователя;
- плохие результаты для простых запросов;

Булева модель, +/−

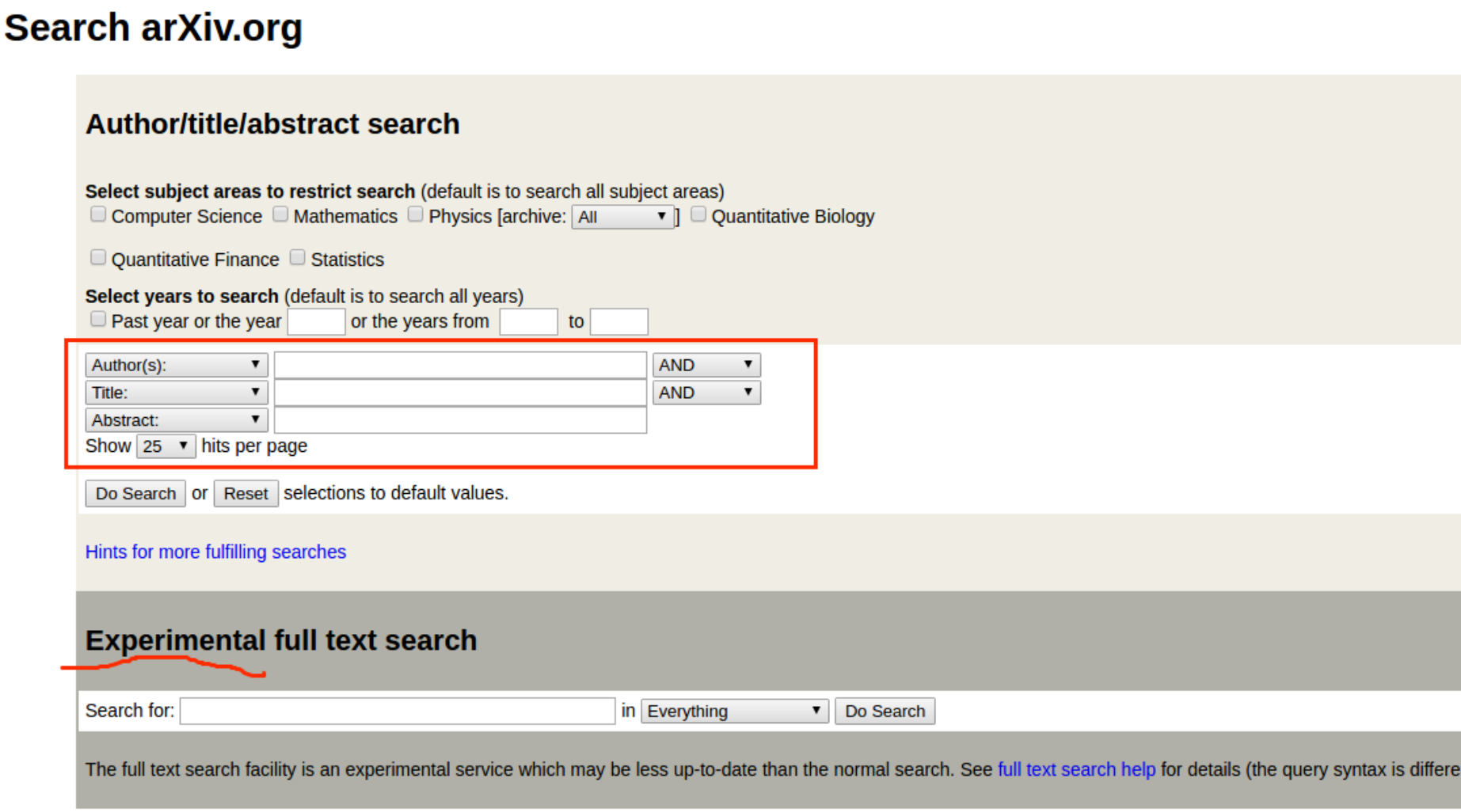
\+

- предсказуемость;
- легкая интерпретация результатов;
- эффективность.

−

- качество полностью зависит от пользователя;
- плохие результаты для простых запросов;
- трудность построения сложных запросов.

Жив ли пациент?



Области применения

- Первая стадия в более сложных поисковых системах.
- Научные электронные библиотеки.
- Правовая сфера.

План

Булев поиск
Инвертированный индекс

Векторная модель

Вероятностные модели в информационном поиске
Языковые модели

Сочетание признаков

Векторная модель поиска

- Запрос задается в произвольной текстовой форме.
- Документы и запросы представлены в виде векторов в T-мерном пространстве, где T – общее количество термов.
- Ранжирование основано на близости векторов в выбранном линейном пространстве.

	1	2	3	4	q
’беда’	0	0	1	0	0
’верста’	0	0	0	1	1
’все’	0	2	0	1	0
’до’	0	0	0	1	0
’за’	0	2	0	0	1
’и’	0	0	0	1	0
’кисель’	0	0	0	0	1
’лес’	0	0	0	1	0
’небеса’	0	0	0	1	0
’один’	1	2	1	0	0
’ответ’	0	0	1	0	0
’отмерить’	1	0	0	0	0
’отрезать’	1	0	0	0	0
’раз’	2	0	0	0	0
’семь’	1	0	1	1	1
’хлебать’	0	0	0	0	1

Ранжирование в векторной модели

$$\text{Score}(\mathbf{q}, \mathbf{d}) = \cos(\vec{q}, \vec{d}) = \frac{(\vec{q}, \vec{d})}{|\vec{q}||\vec{d}|}$$

$$|\vec{x}| = \sqrt{\sum_{t=1}^{T} x_t^2}$$

Взвешивание термов

$$d_t = \text{tf-idf}_{t,d} = TF_{t,d} \times IDF_t$$

$$Score(q, d) = \frac{1}{Z_q \cdot Z_d} \sum_{t|q_t \neq 0} \text{tf-idf}_{t,d} \cdot \text{tf-idf}_{t,q}$$

Z_q, Z_d – нормировочные коэффициенты.

Взвешивание термов

$$d_t = \text{tf-idf}_{t,d} = TF_{t,d} \times IDF_t$$

$$Score(q,d) = \frac{1}{Z_q \cdot Z_d} \sum_{t|q_t \neq 0} \text{tf-idf}_{t,d} \cdot \text{tf-idf}_{t,q}$$

Z_q, Z_d – нормировочные коэффициенты.
для документа:

$$TF_{t,d} = 1 + \log(f_{t,d}), \quad IDF_t = 1, \quad Z_d = \sqrt{\sum_{t=1}^{T} TF_{t,d}^2}$$

для запроса:

$$TF_{t,q} = [f_{t,q} > 0], \quad IDF_t = \log \frac{N}{N_t}, \quad Z_q = \sqrt{\sum_{t=1}^{T} \text{tf-idf}_{t,q}^2}$$

Векторная модель, +/−

+

Векторная модель, +/−

\+

- простота;
- разнообразие вариантов взвешивания термов и мер схожести.

−

Векторная модель, +/−

\+

- простота;
- разнообразие вариантов взвешивания термов и мер схожести.

−

- предположение о независимости термов;
- невозможность определить способ оптимального ранжирования.

План

Булев поиск
Инвертированный индекс

Векторная модель

Вероятностные модели в информационном поиске
Языковые модели

Сочетание признаков

Обоснование

Принцип вероятностного ранжирования (см. предыдущую лекцию).

Постановка задачи

Какова вероятность того, что пользователь оценит данный документ как релевантный для данного запроса?
Нужно оценить $\mathbf{P}(\mathbf{R} = \mathbf{1}|\mathbf{d}, \mathbf{q})$, где $R \in \{0, 1\}$.

Бинарная модель независимости (BIM)

1. Документ и запрос представляются в виде бинарного вектора термов.

$$\vec{x} = (x_1, x_2, \ldots, x_T),\ x_t = [f_{t,x} > 0]$$

2. Термы встречаются независимо друг от друга.

$$P(\vec{x}) = \prod_{t=1}^{T} P(x_t)$$

BIM

По формуле Байеса:

$$P(R=1|\vec{d},\vec{q}) = \frac{P(\vec{d}|R=1,\vec{q})P(R=1|\vec{q})}{P(\vec{d}|\vec{q})}$$

$$P(R=0|\vec{d},\vec{q}) = \frac{P(\vec{d}|R=0,\vec{q})P(R=0|\vec{q})}{P(\vec{d}|\vec{q})} =$$

$$= 1 - P(R=1|\vec{d},\vec{q})$$

BIM

$$O(R|\vec{d},\vec{q}) = \frac{P(R=1|\vec{d},\vec{q})}{P(R=0|\vec{d},\vec{q})} =$$

BIM

$$O(R|\vec{d},\vec{q}) = \frac{P(R=1|\vec{d},\vec{q})}{P(R=0|\vec{d},\vec{q})} =$$

$$= \frac{P(\vec{d}|R=1,\vec{q})}{P(\vec{d}|R=0,\vec{q})} \cdot \frac{P(R=1|\vec{q})}{P(R=0|\vec{q})} \overset{(2)}{=} O(R|\vec{q}) \cdot \prod_{t=1}^{T} \frac{P(d_t|R=1,\vec{q})}{P(d_t|R=0,\vec{q})} \overset{(1)}{=}$$

$$\overset{(1)}{=} O(R|\vec{q}) \cdot \prod_{t:d_t=1}^{T} \frac{P(d_t=1|R=1,\vec{q})}{P(d_t=1|R=0,\vec{q})} \prod_{t:d_t=0}^{T} \frac{P(d_t=0|R=1,\vec{q})}{P(d_t=0|R=0,\vec{q})} =$$

$$= O(R|\vec{q}) \cdot \prod_{t:d_t=1}^{T} \frac{\mathbf{p_t}}{\mathbf{u_t}} \prod_{t:d_t=0}^{T} \frac{\mathbf{1-p_t}}{\mathbf{1-u_t}}$$

BIM

Предположим, что если $q_t = 0$, то $p_t = u_t$.

$$O(R|\vec{d},\vec{q}) = O(R|\vec{q}) \cdot \prod_{t:d_t=q_t=1} \frac{p_t}{u_t} \prod_{t:d_t=0,q_t=1} \frac{1-p_t}{1-u_t} =$$

$$= O(R|\vec{q}) \cdot \prod_{t:d_t=q_t=1} \frac{p_t(1-u_t)}{u_t(1-p_t)} \cdot \prod_{t:q_t=1} \frac{1-p_t}{1-u_t} \quad (1)$$

BIM

Предположим, что если $q_t = 0$, то $p_t = u_t$.

$$O(R|\vec{d},\vec{q}) = O(R|\vec{q}) \cdot \prod_{t:d_t=q_t=1}^{T} \frac{p_t}{u_t} \prod_{t:d_t=0,q_t=1}^{T} \frac{1-p_t}{1-u_t} =$$

$$= O(R|\vec{q}) \cdot \prod_{t:d_t=q_t=1}^{T} \frac{p_t(1-u_t)}{u_t(1-p_t)} \cdot \prod_{t:q_t=1}^{T} \frac{1-p_t}{1-u_t} \quad (1)$$

Retrieval Status Value

$$RSV_d = \sum_{t:d_t=q_t=1} \log \frac{p_t(1-u_t)}{(1-p_t)u_t} = \sum_{t:d_t=q_t=1} c_t \quad (2)$$

Оценка вероятностей

S – количество релевантных запросу q документов в коллекции, S_t – количество релевантных, содержащих терм t.

$$p_t = \frac{S_t}{S}, \quad u_t = \frac{N_t - S_t}{N - S}$$

Оценка вероятностей

S – количество релевантных запросу q документов в коллекции, S_t – количество релевантных, содержащих терм t.

$$p_t = \frac{S_t}{S}, \quad u_t = \frac{N_t - S_t}{N - S}$$

$$c_t = \log \frac{S_t/(S - S_t)}{(N_t - S_t)/(N - N_t - S + S_t)} \approx$$
$$\approx \log \frac{(S_t + 0.5)/(S - S_t + 0.5)}{(N_t - S_t + 0.5)/(N - N_t - S + S_t + 0.5)} \quad (3)$$

Небинарная модель: Okapi BM25

$$RSV_d = \sum_{t:q_t=1} \left(\log \frac{(S_t + 0.5)/(S - S_t + 0.5)}{(N_t - S_t + 0.5)/(N - N_t - S + S_t + 0.5)} \times \right.$$
$$\left. \times \frac{(\mathbf{k_1} + 1) f_{t,d}}{\mathbf{k_1}((1 - \mathbf{b}) + \mathbf{b} \cdot L_d/\bar{L}) + f_{t,d}} \times \frac{(\mathbf{k_2} + 1) f_{t,q}}{\mathbf{k_2} + f_{t,q}} \right)$$

$b = 0.75$, $k_1 = 1.2$, $k_2 = 0..1000$
L_d – длина документа d, $\bar{L}$ – средняя длина документа в коллекции.

Снова TF-IDF

$$S_t = S = 0,\ \log \frac{N - N_t}{N_t} \approx \log \frac{N}{N_t}$$

$$RSV_d =$$

$$= \sum_{t:q_t=1} \left(\log \frac{N}{N_t} \times \frac{(k_1+1) f_{t,d}}{k_1((1-b) + b \cdot L_d/\bar{L}) + f_{t,d}} \times \frac{(k_2+1) f_{t,q}}{k_2 + f_{t,q}} \right) =$$

$$= \sum_{t:q_t=1} IDF_t \times TF_{t,d} \times TF_{t,q}$$

Ранжирование с использованием языкового моделирования

Языковая модель – функция, приписывающая каждой строке над некоторым словарем некоторую вероятность.

$$P(R = 1|q, d) \approx P(d|q) = \frac{P(q|M_d)P(d)}{P(q)}$$

M_d – языковая модель документа d.

Униграммная модель

$$P(\vec{q}|M_d) = \prod_{t:q_t=1} P(q_t|M_d) = \prod_{t:q_t=1} \frac{f_{t,d}}{L_d}$$

План

Булев поиск
Инвертированный индекс

Векторная модель

Вероятностные модели в информационном поиске
Языковые модели

Сочетание признаков

Learning to Rank

$$Score(q, d) = F(f_1, f_2, \ldots, f_k)$$

, где $f_i = f_i(q, d)$ – признак.
$F \in \mathcal{F}$ подбирается на обучающей выборке $\{(f_{i1}, \ldots, f_{ik}; y_i)\}_{i=1}^{n}$ исходя из задачи оптимизации некоторой метрики.
y_i – оценки релевантности, выраженные в ассессорских оценках либо пользовательских действиях.

Learning to Rank

- вычисление таких моделей достаточно тяжелая операция, поэтому применяются к ограниченному числу документов;
- значимые финансовые затраты для составления обучающей выборки;
- тем не менее, зависят от качества более низкоуровневых моделей.

Следующая лекция

Компьютерная лингвистика в информационном поиске